

TITLE OF THE INVENTION

CONFIGURABLE SPEECH RECOGNIZER

FIELD OF THE INVENTION

[0001] The present invention relates to speech recognition generally and to configurable speech recognizers in particular.

BACKGROUND OF THE INVENTION

[0002] Speech recognition is known in the art. Limited vocabulary speech recognizers operate by matching the incoming speech to a collection of reference speech models and selecting the reference model(s) which best match(es) the incoming speech. Limited vocabulary speech recognizers are used for speech dialing, in which the user says a phone number and the speech recognizer determines which digits were said and provides the recognized digits to the automatic dialing system of a telephone. "Digits" typically include the numerical digits, symbols, such as *, # and +, and pause and editing words such as "clear", "cancel", "dial" and save". Speech dialers exist on cellular telephones to provide 'hands-free dialing' during driving.

[0003] Speech dialers, especially those operating in a car environment, often have difficulty determining the digits said, since some digits have similar sounding names in certain languages. To improve recognition performance, some speech recognition systems add constraints to the recognition process, based on the natural constraints of the dialing process.

[0004] For speech dialing, if the user defines the country where the phone is used, then the "numbering plan" of that country may be used to constrain at least some of the digits. For

example, the numbering plan of the United States states that the first number of an area code may not be a 0 or a 1. Furthermore, all area codes are comprised of three digits, all exchanges are comprised of three digits, and there are four remaining digits. A more complete numbering plan for the US is listed below, where N is a digit from 2 – 9 and X is a digit from 0 – 9 and ‘-’ indicates an end of a phrase:

1 digit number: 0 (operator)

3 digit number: N11

3 digit number: *XX

4 digit number: *XXX

7 digit number: NXX-XXXX

10 digit number: NXX-NXX-XXXX

11 digit number: 1-NXX-NXX-XXXX

11 digit number: 0-NXX-NXX-XXXX

[0005] If the user says three digits, then the speech dialer, using the numbering plan, can ‘guess’ that the first digit was either a star (*) or an N. Similarly, if seven digits were said, then the first digit cannot be a zero or a one. This slight constraining of the speech recognizer significantly improves the recognition results. In addition to the hard constraints described above, speech recognizers sometimes apply soft constraints, i.e. all digit sequences are allowed but prior probabilities are used to elevate the probabilities of recognizing certain sequences and reduce the probabilities of others.

[0006] In 2003, the following websites included in them descriptions of various numbering plans:

- [0007] World Telephone Numbering Guide: [www. wtng. info](http://www.wtng.info)
- [0008] North American Numbering Plan Administration: [www. nanpa. com](http://www.nanpa.com)
- [0009] Vertical service codes (dialing numbers specific to carriers):
- [0010] [www. nanpa. com/number_resource_info/vertical_service.html](http://www.nanpa.com/number_resource_info/vertical_service.html)
- [0011] [www. nanpa. com/number_resource_info/vsc_assignments.html](http://www.nanpa.com/number_resource_info/vsc_assignments.html)

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] The subject matter regarded as the invention is particularly pointed out and distinctly claimed in the concluding portion of the specification. The invention, however, both as to organization and method of operation, together with objects, features, and advantages thereof, may best be understood by reference to the following detailed description when read with the accompanying drawings in which:

[0013] Fig. 1 is a block diagram illustration of an exemplary part of cellular telephone, constructed and operative in accordance with the present invention; and

[0014] Fig. 2 is a block diagram illustration of a configurable speech recognizer forming part of the telephone of Fig. 1.

[0015] It will be appreciated that, where considered appropriate, reference numerals may be repeated among the figures to indicate corresponding or analogous elements.

DETAILED DESCRIPTION OF THE INVENTION

[0016] In the following detailed description, numerous specific details are set forth in order to provide a thorough understanding of the invention. However, it will be understood by those skilled in the art that the present invention may be practiced without these specific details. In other instances, well-known methods, procedures and components have not been described in detail so as not to obscure the present invention.

[0017] Cell phones receive geographic information in many forms. Applicants have realized that some of the geographic information may be transferred to a speech recognition system to help it automatically configure the portions of its associated search information which are a function of geography. For example, the associated search information may include the numbering plans of each country and/or dialect information, etc. In the present invention, when the cell phone determines that the user has changed countries, it may pass this information to the speech recognition system which, in turn, may select the appropriate numbering plan for that country. Other geographically related search information, such as dialects, preferred pronunciations, etc., may also be selected.

[0018] The geographic information to be used may be of many forms. It may be the location of the current base station with which the cell phone is communicating. In another example, the geographic information may be taken from the operator identification number transmitted when a cell phone may start up or when it may be “roaming”, since such includes regional information in it.

[0019] Briefly, roaming is the ability of a system to provide the same services to customers (‘roamer’) from other systems. It is discussed, in 2003, in the following websites:

[0020] www.gsmworld.com/roaming/gsminfo/index.shtml

[0021] www.cdg.org/technology/roaming.asp

[0022] In addition, in the present invention, geographic information may also include operator specific information, such as telephone dialing styles specific to an operator.

[0023] Reference is now made to Fig. 1, which illustrates a cellular telephone 10, constructed and operative in accordance with the present invention. Telephone 10 may comprise a cellular telephony unit 12, a geographic location determiner 14 and a configurable speech dialer 16 and may use geographic or operator information to limit the search space of dialer 16. Cellular telephony unit 12 may be the portion of a cellular telephone which may provide the standard cellular telephony services, including the ability to roam from one cellular telephone operator to another. Roaming may also occur when cellular telephony unit 12 may leave the cellular network and may become an extension of a landline system. Such may occur with dual mode GSM/DECT phones. While the user is out of the office or away from home the phone may communicate with the wide-area GSM cellular network. While the user is at home or in the office, the phone may communicate via a wireless DECT base station. Such a phone is known as "one phone anywhere" and is described at [http:// www.dectweb.com/Products/dual_mode.htm](http://www.dectweb.com/Products/dual_mode.htm) in 2003.

[0024] Configurable speech dialer 16 may be any speech dialer which may have a multiplicity of constraints therein and which may change these constraints when provided with a configuration signal. An exemplary dialer 16 is shown in Fig. 2, discussed hereinbelow.

[0025] When telephone 12 may start up, or when it may roam from one operator area to another, cellular telephony unit 12 may provide geographic location determiner 14 with the geographic (or roaming) information which, in turn, may process it to retrieve the appropriate geographic information. For speech dialer 16, the appropriate geographic information may be the country or region of a country where the cellular telephone operator may be located. Geographic location determiner 14 may pass the country information to speech dialer 16 which may reconfigure itself to use the numbering plan of the new country or region. The numbering plan may also change between cellular and landline operators since many operators may have some additional numbering styles of their own.

[0026] Geographic location determiner 14 may process the roaming information to find the portion of it which provides geographic information. Determiner 20 may translate this information into a country or regional identification, or a location indication (home/office vs. external), and may pass this identification to speech dialer 16.

[0027] It will be appreciated that the present invention may utilize any geographic information that cellular telephony unit 12 may have. This information may include, but is not limited to, the cellular operator identification number, other roaming information, any GPS information that the cellular telephone may generate, and location information that the cellular telephone may generate from the cellular network.

[0028] Reference is now made to Fig. 2 which illustrates an exemplary configurable speech dialer 16. Dialer 16 may comprise a search engine 30, a recognition manager 31, a multiplicity of knowledge bases 32 and a multiplicity of reference libraries 34.

[0029] Each knowledge base 32 may contain information about the expected speech patterns for one geographic area. This information may include accents, dialects, preferences for particular words, etc. For example, the preferred way to pronounce the symbol # is "pound" in the United States, but "hash" in the United Kingdom. Likewise, the expected grouping of digits or placement of pauses through the utterance varies according to geographical region; in Canada this may be grouped as 3 digits, 3 digits, then 4 digits, while in France it may be 5 groups of 2 digit numbers. Similarly, the expected way of entering the phone number varies; it may be as one utterance of all the digits as seen in the Motorola Spirit car phone, or as variable size groups of digits and editing commands, as demonstrated on the Siemens Xelibri 3 phone.

[0030] Each knowledge base 32 may also contain personalized information, such as the list of the latest dialed phone numbers. Knowledge base 32 may also be updated with the operator or company numbering plan (such as vertical numbers or internal extensions). In addition, for speech dialing, each knowledge base 32 may contain the numbering plan of the geographic area. Knowledge base 32 may either contain pre-stored operator specific numbering plans or a current plan to be used may be transmitted to the phone from the operator. The operator specific numbering plan may contain short-dialing options and vertical service codes specific to the operator. Update of knowledge base 32 may be performed during the manufacturing of the phone, software installation by the operator, and over the air.

[0031] Each reference library 34 may contain a set of acoustic models representative of a specific language or regional dialect. Reference libraries 34 may also contain acoustic models

representative of different words, according to the preferred way of speaking phone numbers in the geographical area.

[0032] Recognition manager 31 may receive the location information from geographic location determiner 14 (Fig. 1) and may select the knowledge base 32A associated with the country, region, cellular telephone operator and/or company information indicated by the location information.

[0033] Recognition manager 31 may then supply search engine 30 with the appropriate reference library 34A according to information from the active knowledge base 32A about one or more of: language, accent, dialect and/or region specific words. For example, recognition manager 31 may select active reference library 34A according to regional dialect while various acoustic models within library 34A may be selected according to the preferred pronunciation of various digits and symbols for the location, the expected way of entering phone numbers, and the other information described hereinabove.

[0034] Recognition manager 31 may also set the grammar to be used by the search engine according to information from active knowledge base 32A about the numbering plan and/or placement of pauses and/or last dialed calls.

[0035] Search engine 30 may attempt to match an incoming speech signal with a set of reference models, such as HMM or template models, stored in active reference library 34A, producing the digits to be dialed as output. Search engine 30 may utilize the information in active knowledge base 32A to constrain the number and type of reference models (from active reference library 34A) to which the input speech signal may be matched using the grammar provided by recognition manager 31.

[0036] In another embodiment, search engine 30 may apply soft constraints according to the operator information, the geographic location and/or the numbering plan. In this embodiment, non-numbering plan numbers may not be blocked; however, the recognition of numbering plan numbers may be improved.

[0037] In addition, during an editing mode, the current grammar may be changed after each utterance, according to the remaining valid rules. For example, if the allowed numbering plan in the United States is 7 or 10 digits and the speaker has already uttered 3 digits, the grammar may be changed to expect 4 or 7 digits.

[0038] It will further be appreciated that speech dialing is only one application of the present invention. There are many other speech recognition operations which may be usefully constrained with geographic information. For example, accents, dialects and vocabulary all vary from one region to another. In another example, some speech recognizers may have to recognize the names of locations. Knowing the general region where a telephone, which will be transmitting the name of a location, is may help to constrain the search space.

[0039] For phoneme-based speech recognition tasks, such as name dialing according to the text written in the phonebook, recognition manager 31 may also use the geographical information to set the text to phoneme conversion module of search engine 30. For example, the expected pronunciation of French names in English speaking regions of Canada may be different than the pronunciation in French speaking regions. A geographical cue may be used in this case to introduce prior probability to the text to phoneme module. Additionally, different transcription libraries can be used according to the geographical location.

[0040] As can be seen, any speech recognition task which may have some constraints which are geographically related may utilize the present invention.

[0041] While certain features of the invention have been illustrated and described herein, many modifications, substitutions, changes, and equivalents will now occur to those of ordinary skill in the art. It is, therefore, to be understood that the appended claims are intended to cover all such modifications and changes as fall within the true spirit of the invention.